

# Modelos Não Paramétricos

Professor: Pedro M. Almeida-Junior

---

5 de junho de 2021

Departamento de Estatística (UEPB)

# Regressão Não paramétrica

---

# Introdução

- Uma curva de regressão descreve uma relação geral entre uma variável explicativa  $X$  e uma variável de resposta  $Y$ .
- Tendo observado  $X$ , o valor médio de  $Y$  é dado em termos de uma função de regressão.
- Se  $n$  pontos de dados  $\{(X_i, Y_i)\}_{i=1}^n$  foram coletados, a relação de regressão pode ser modelado como

$$Y_i = m(X_i) + \epsilon_i, \quad i = 1, \dots, n$$

com a função de regressão desconhecida  $m$  e erros de observação  $\epsilon_i$ .

- O objetivo de uma análise de regressão é produzir uma análise razoável para a função de resposta desconhecida  $m$ .
- Ao reduzir os erros observacionais, permite que a interpretação se concentre em detalhes importantes da dependência média de  $Y$  em  $X$ . Este procedimento de aproximação de curva é comumente chamado de "suavização".
- Essa tarefa de aproximar a função média pode ser feita essencialmente de duas maneiras: (i) Abordagem paramétrica e (ii) Abordagem não paramétrica.

- A abordagem paramétrica frequentemente usada é assumir que a curva média  $m$  tem alguma forma funcional pré-especificada, por exemplo, uma linha com inclinação e interceptação desconhecidas.
- A forma funcional é totalmente descrita por um conjunto finito de parâmetros.
- O termo não paramétrico, refere-se à forma funcional flexível da curva de regressão. Existem outras noções de "estatísticas não paramétricas" que se referem principalmente a métodos livres de distribuição.
- No presente contexto, entretanto, nem a distribuição do erro nem a forma funcional da função média são pré-especificadas.

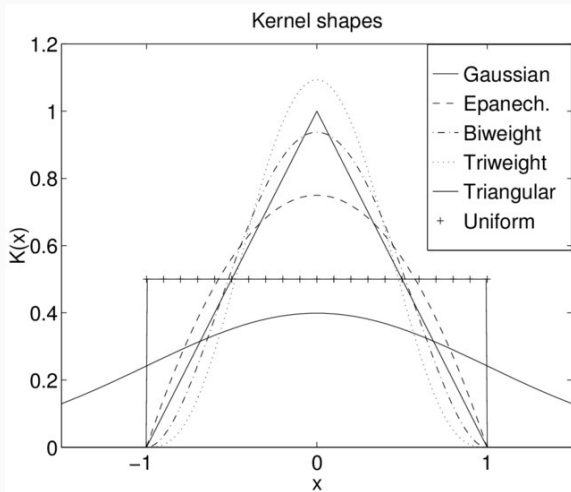
# Objetivos Abordagem Não Paramétrica

- A abordagem não paramétrica para estimar uma curva de regressão tem como objetivos principais:
  1. Fornecer um método versátil de explorar uma relação geral entre duas variáveis.
  2. Fornecer previsões de observações ainda a serem feitas sem referência a um modelo paramétrico fixo.

- Uma abordagem conceitualmente simples para uma representação da sequência de peso  $\{W_{ni}(x)\}_{i=1}^n$  é descrever a forma da função de peso  $W_{ni}(x)$  por uma função de densidade com um parâmetro de escala que ajusta o tamanho e a forma dos pesos próximos a  $x$ .
- É bastante comum referir-se a essa função de forma como um kernel  $K$ . O kernel é uma função real contínua, limitada e simétrica  $K$  que se integra a um,

$$\int K(u)du = 1$$

# Tipos de Kernels



**Figura 1:** Tipos de Kernels



A sequência de peso para suavizadores de kernel (para  $x$  unidimensional) é definida por

$$W_{ni}(x) = K_h(x - x_i) / \hat{f}_h(x),$$

em que

$$\hat{f}_h(x) = n^{-1} \sum_{i=1}^n K_h(x - x_i)$$

e

$$K_h(u) = h^{-1} K(u/h)$$

## Algoritmo de classificação $k$ -NN

---

# Algoritmo $k$ -NN

O algoritmo  $k$ -Nearest Neighbors ( $k$ -NN) foi proposto por Cover e Hart e é considerado um classificador não paramétrico. O algoritmo  $k$ -NN possui os seguintes recursos:

- ✓  $k$ -NN é um algoritmo de aprendizado supervisionado que usa um conjunto de dados de entrada rotulado para prever a saída dos pontos de dados.
- ✓ É um dos algoritmos de aprendizado de máquina simples e pode ser facilmente implementado para um conjunto variado de problemas.
- ✓  $k$ -NN verifica a semelhança de um ponto de dados com seu vizinho e classifica o ponto de dados na classe com a qual é mais semelhante.

- Dada uma amostra de teste onde a classe é desconhecida, o método encontra  $k$  mais próximo no conjunto de treinamento para cada observação da amostra de teste segundo uma determinada distância (geralmente utilizada distância euclideana) e atribui uma classe à observação de acordo com o voto da maioria das classes desses vizinhos;
- A distância euclidiana entre os pontos  $P = (p_1, p_2, \dots, p_n)$  e  $Q = (q_1, q_2, \dots, q_n)$ , num espaço euclidiano  $n$ -dimensional, é definida como:

$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

# Distância de Mahalanobis

Formalmente, a distância de Mahalanobis entre um grupo de valores com média  $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3, \dots, \mu_p)^T$  e matriz de covariância  $S$  para um vetor multivariado  $\mathbf{x} = (x_1, x_2, x_3, \dots, x_p)^T$  é definida como:

$$D_M(x) = \sqrt{(\mathbf{x} - \boldsymbol{\mu})^T S^{-1} (\mathbf{x} - \boldsymbol{\mu})}$$

- O algoritmo estima a probabilidade de uma observação pertencer a cada grupo com base na informação do vizinho mais próximo  $k$  definida a seguir.
- Considere  $G$  partições não vazias, que são  $C_1, C_2, \dots, C_G$ , tal que,  $C_j \subset \mathbb{R}^p$  é um subconjunto de tamanho  $n_j$ , para  $j = 1, \dots, g$ ,  $n = \sum_{j=1}^n n_j$ ,  $C_i \cap C_j = \emptyset$  para todo  $i \neq j$ .

- Seja  $\mathbf{y} \in \mathbb{R}^p$  uma observação a ser classificada e  $\{\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_k}\}$  o conjunto de  $k$  vizinhos, tal que,  $k \leq n$  e  $\{i_1, \dots, i_k\} \subset \{1, \dots, n\}$ . Assim para  $j = 1, \dots, G$ , a regra de classificação de  $k$ -NN é definida como:

$$\hat{c}_{kNN} = \arg \max_c \hat{P}_k(G = c)$$

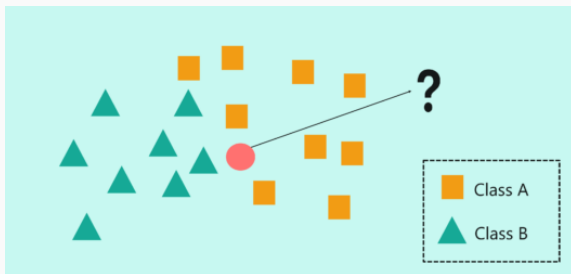
em que

$$\hat{P}_k(G = c) = k^{-1} \sum_{\nu=1}^k \mathbb{I}_c(\mathbf{x}_{i_\nu})$$

e

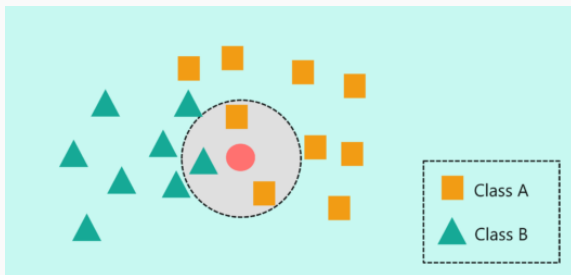
$$\mathbb{I}_c(\mathbf{x}_{i_\nu}) = \begin{cases} 1, & \text{if } \mathbf{x}_{i_\nu} \in C_j \\ 0, & \text{if } \mathbf{x}_{i_\nu} \notin C_j \end{cases} \quad (1)$$

# Ilustração Algoritmo $k$ -NN

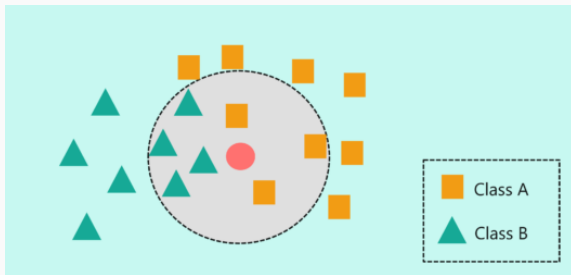


**Figura 2:** Ilustração  $k$ -NN





**Figura 3:**  $k$ -NN com  $k = 3$  vizinhos



**Figura 4:**  $k$ -NN com  $k = 7$  vizinhos